Lost in Magnitudes: Exploring Visualization Designs for Large Value Ranges

Katerina Batziakoudi, Florent Cabric, Stéphanie Rey, Jean-Daniel Fekete

▶ **To cite this version:**

Katerina Batziakoudi, Florent Cabric, Stéphanie Rey, Jean-Daniel Fekete. Lost in Magnitudes: Exploring Visualization Designs for Large Value Ranges. CHI 2025 - CHI Conference on Human Factors in Computing Systems, ACM, Apr 2025, Yokohama, Japan. pp.1-18 Article No.: 1170, 10.1145/3706598.3713487 . hal-04981028

HAL Id: hal-04981028

https://inria.hal.science/hal-04981028v1

Submitted on 6 Mar 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

Lost in Magnitudes: Exploring Visualization Designs for Large Value Ranges

Katerina Batziakoudi
Berger-Levrault, Université Paris-Saclay, CNRS, Inria
Paris, France
a.batziakoudi@berger-levrault.com

Florent Cabric
Université Paris-Saclay, CNRS, Inria
Paris, France
florent.cabric@inria.fr

Stéphanie Rey
Berger-Levrault
Toulouse, France
Stephanie.REY@berger-levrault.com

Jean-Daniel Fekete
Université Paris-Saclay, CNRS, Inria
Paris, France
Jean-Daniel.Fekete@inria.fr

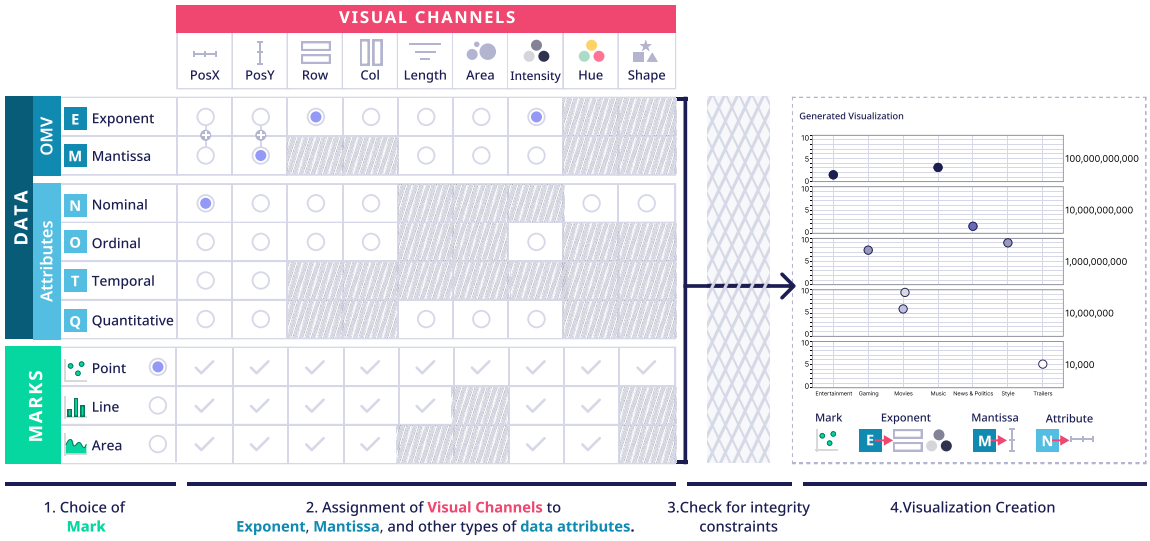

Figure 1: Our design space has three dimensions: MARKS (green), DATA (blue), and VISUAL CHANNELS (red). The image shows an example of using our design space as an interactive table, where a mark is selected, and visual channels are assigned to data attributes. Grey cells are invalid, according to the visualization literature. After checking for integrity constraints, a visualization is generated. Note the ⊕ signs (top left); enable the assignment of the `EplusM` scale to positions.

ABSTRACT

We explore the design of visualizations for values spanning multiple orders of magnitude; we call them Orders of Magnitude Values (OMVs). Visualization researchers have shown that separating OMVs into two components, the *mantissa* and the *exponent*, and encoding them separately overcomes limitations of linear and logarithmic scales. However, only a small number of such visualizations have been tested, and the design guidelines for visualizing the mantissa and exponent separately remain under-explored. To initiate this exploration, better understand the factors influencing the effectiveness of these visualizations, and create guidelines, we adopt a multi-stage workflow. We introduce a design space for visualizing mantissa and exponent, systematically generating and qualitatively evaluating all possible visualizations within it. From this evaluation, we derive guidelines. We select two visualizations that align with our guidelines and test them using a crowdsourcing experiment, showing they facilitate quantitative comparisons and increase confidence in interpretation compared to the state-of-the-art.

CCS CONCEPTS

• **Human-centered computing** → **Visualization theory, concepts and paradigms**.

KEYWORDS

design space, visualization, static, overview, orders of magnitude, mantissa, exponent

Please use nonacm option or ACM Engage class to enable CC licenses
This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.
CHI '25, April 26-May 1, 2025, Yokohama, Japan
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713487

ACM Reference Format:
Katerina Batziakoudi, Florent Cabric, Stéphanie Rey, and Jean-Daniel Fekete. 2025. Lost in Magnitudes: Exploring Visualization Designs for Large Value Ranges. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26-May 1, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3706598.3713487

1 INTRODUCTION

In this article, we explore the design of static visualizations with values spanning multiple orders of magnitude, drawing inspiration from scientific notation and floating-point arithmetic that separate values into an exponent and a mantissa part, such that a value v is expressed as $v = mantissa \times 10^{exponent}$. Order of Magnitude Values (OMVs) are integral to various domains of daily life, including but not limited to financial analysis, pandemic tracking, demographic studies, environmental monitoring, and social media metrics. An illustrative example is the budget allocations of the French government (Figure 2), which range from tens of millions of Euros (10^7€) to hundreds of billions of Euros (10^{11}€), thereby covering five orders of magnitude. The effective static visualization of OMVs is essential for providing an informative overview suitable for both printed and digital media.

Common visualization techniques, such as linear and logarithmic bar charts, prove ineffective in visualizing OMVs. In the example of the French budget (Figure 2), using a linear scale (left) makes it impossible to read and compare budget categories with smaller magnitudes, which are shown with less than one pixel, resulting in information being entirely lost. While a logarithmic scale (right) can partially mitigate this issue by ensuring that all budget values are visible, it is not well understood by the general public [17, 48, 58] and may even lead to misconceptions among scientists [38]. Beyond legibility issues, logarithmic scales also hinder efficient quantitative comparisons [29], such as the low-level tasks of estimating ratios between values within the same or neighboring magnitudes.

Visualization researchers, also inspired by scientific notation, have proposed novel static designs that address the limitations of linear and logarithmic scales [6, 10, 11, 29, 30] (see Figure 3). They separate OMVs into two parts: the *mantissa* and the *exponent* and visualize each component separately using different visual channels [10, 11, 29, 30], distinct marks [6], or novel scales [10, 11, 29]. This decomposition into two parts challenges established graphical perception heuristics for magnitude estimation, as a single quantitative attribute is represented by two interdependent components. Low-level tasks, such as value retrieval and comparison, require the original OMV to be reconstructed by visually combining the mantissa and exponent, introducing novel constraints and requirements. While researchers have empirically demonstrated that this separation is beneficial for visualizing OMVs, only a few visualizations have been explored so far, and the design principles for effectively visualizing mantissa and exponent remain unclear.

Building on this foundation, to better understand the factors affecting the effectiveness of visualizations that separate OMVs to mantissa and exponent we follow a multistage workflow. We begin by **reviewing existing literature** on large number estimation, graphical perception, and visualization of large value ranges to **define design requirements and constraints**. Next, we **describe a design space**, rooted in the principles of the Grammar

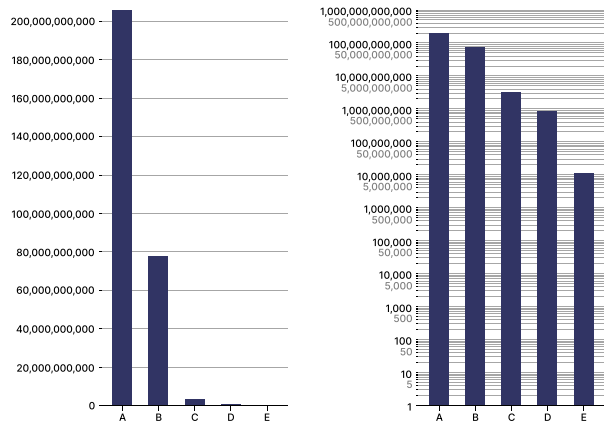

Figure 2: Linear bar chart (left) and logarithmic bar chart (right) illustrating a sample of the French government's budget allocations, showcasing order-of-magnitude differences.

of Graphics (GoG) [74] and focused on tabular data, the Cartesian system, and visualizations with no redundant encoding. We **refine the design space** by applying constraints derived from the literature. We **systematically generate** all possible visualizations within this defined space and we **qualitatively evaluate** them, aiming to identify encoding and decoding problems that could impact their effectiveness. Building on the results of the qualitative evaluation, we **derive design guidelines** and **test them through a crowdsoursing experiment** comparing two newly designed visualizations, aligned with these guidelines, against traditional linear and logarithmic bar charts, as well as a visualization from the state-of-the-art (the Scale-Stacked Bar Chart). Our results demonstrate that the proposed designs performed similarly to, or better than, existing solutions and confirmed that visualizations that both respect our guidelines and separate OMVs into mantissa and exponent facilitate quantitative comparisons.

In summary, we make the following contributions:

- we present a design space for the static visualization of OMVs, when mantissa and exponent are separately visualized
- we introduce guidelines for designing effective mantissa-exponent visualizations for OMVs
- we introduce and validate two visualization designs for OMVs, Facet and `EplusM`, that are effective for quantitative tasks.

Additionally, we address error measurements in OMV experiments to improve the comparability of results in future studies. We also open-sourced the tool developed for our design space exploration, along with all the generated visualizations, to encourage further evaluation and expansion of the defined design space, as well as to support the application of this methodology to other design space exploration challenges. The numerical separation of OMVs into mantissa and exponent has long been valued for its advantages in computational efficiency, numerical stability, and readability across scientific disciplines. Inspired by these strengths, our work demonstrates how this separation can benefit the visualization domain, paving the way for broader adoption and generalization of mantissa-exponent visualizations for OMVs.

2 BACKGROUND

In this section, we first discuss the definition and numerical relationship between the mantissa and exponent. Then, we present related work in data visualization with varying orders of magnitude and visualization effectiveness.

2.1 Definition and Terminology

We borrow our definitions and terminology from both *scientific notation*—a notation for symbolically writing OMVs—and floating-point numbers (see [36, section 4.2] for a formal introduction to floating point numbers)—a computer representation of OMVs used to perform calculations. A numerical value v is represented as a pair (m, e) with $v = m \times b^e$, where $m \in \mathbb{R} \setminus \{0\}$ is called *mantissa* (sometimes *significand* or *fraction*) and $e \in \mathbb{Z}$ is called *exponent*. Like scientific notation, we use *base* $b = 10$, whereas floating-point numbers use base 2.

In *scientific normalized notation*, e is chosen so that $|m| \in [1, 10)$. The mantissa defines the *significant digits* of a value inside an order of magnitude and is represented with a given *precision*, the number of digits $p \in \mathbb{N}^+$, that may depend on v's intrinsic precision and on the desired/available display precision.

In accordance with Höhn et al. [30], **a dataset contains orders of magnitude values (OMVs) (or a *large value range*) when one of its numerical attributes contains values that span four orders of magnitude or more.** Using scientific notation and floating point numbers allows for dealing with large ranges of values but also introduces new issues compared to decimal notation [25]. In particular, regarding visualization of values spanning multiple orders of magnitude (hereafter referred to as *OMVs visualization*) , the order of magnitude of 0 is undefined, and a smaller number can be *negligible* compared to a larger one when their precision is lower than the difference between their exponents.

2.2 Visualization and Perception of Large Value Ranges

In the context of data visualization, visually representing OMVs poses significant challenges. While linear scales are widely understood, they fall short of effectively depicting large value ranges, rendering only the larger values legibly. If the screen has a width and height in the 10^3 pixels, linear scale is not capable of visualizing values that are more than three orders of magnitude apart, as shown in Figure 2. This distinctiveness issue [56] hinders fundamental analytical tasks, including value estimation and comparison (ratio and difference) between different ranges. Although dual scale charts [31] and innovative designs like the Du Bois Wrapped Bar Chart [34] offer improvements for analyzing two disparate ranges, their effectiveness diminishes with OMVs [6], as they require the accurate representations of multiple (i.e., more than four) data ranges in the same view.

Logarithmic scales, capable of showcasing diverse ranges, facilitate some qualitative tasks with OMVs, like finding extrema [30]. Nonetheless, they pose difficulties for quantitative tasks, such as estimating differences or ratios between values of similar or close magnitudes [29]. Despite their prevalence in scientific contexts, they may lead to misconceptions [38] and are less understandable for the general public, challenging perception and interpretation [17, 40, 59]. Research results in cognitive psychology [17, 21, 39, 40] showcase a common misconception: individuals frequently interpret the spacing between major ticks on a logarithmic scale linearly. This segmented or piecewise linear interpretation becomes particularly pronounced when dealing with values exceeding one million, aligning with how large magnitudes are verbally and symbolically represented [39, 40].

When describing large numbers verbally, we use a hybrid notation comprising two components: a number ranging between 1 and 999 to indicate precision and a scale-word (e.g., millions, billions) to represent magnitude. Similarly, when we write large numbers using scientific notation we use two components: the mantissa for precision and the exponent for magnitude. Both notations rely on a structured schema that separates a value into two parts, linked by a mathematical relationship. The exponent increases multiplicatively (e.g., powers of 10), while the mantissa increases additively, showing a linear progression within the range defined by two exponents. In the following sections we use the term *continuity* to refer to this mathematical relationship between mantissa and exponent.

In the past decade, researchers have developed visualizations that separate OMVs into exponent and mantissa components, enhancing tasks like value retrieval, sorting, ratio estimation, difference estimation, and trend identification. Hlawatsch et al. [29] introduced the Scale-Stack Bar Chart (**SSB**) that visualizes OMVs by stacking values at multiple scales, beginning from zero and mapping numbers linearly. Borgo et al. [6] visualized exponents and mantissas as overlapping bars, termed Order of Magnitude Markers (**OMM**). Höhn et al. [30] presented Width-Scale Bar Charts (**WSB**), using Y position for mantissa and color plus width for the exponent. Additionally, Braun et al. [11] used color in their Order of Magnitude Color (**OMC**) scheme to map exponents to hue and mantissa to a color scale. Focusing on large-scale time series, Braun et al. [10] introduced the Order of Magnitude Horizon Graph (**OMH**) and Line Chart (**OML**), with the first using Y position for mantissa and color intensity for exponents, and the second combining the positional encoding from SSB and the OMC color scale. Unlike SSB, where values are shown per magnitude in stacked rows, OML use "linear scaling within orders of magnitude" combining both exponent and mantissa in the Y position [10]. Figure 3 shows examples of these visualizations, and Table 1 summarises existing literature.

The studies highlighted above have conclusively demonstrated some advantages of visualizations that separate OMVs into their mantissa and exponent components. This separation enhances analytical tasks' accuracy, confidence, and efficiency by offering a more effective representation of OMVs. However, these studies have explored only a subset of the potential combinations of marks and visual channels for encoding the mantissa and exponent (Table 1). Although this subset is promising, there is still a lack of a systematic approach in understanding the factors affecting the effectiveness of mantissa and exponent visualizations. To address this gap, our objective is to define a design space for visualizing mantissa and exponent, systematically explore it, document observed problems that may inform design guidelines, and identify effective encodings, enriching the corpus of OMV visualizations. We base our exploration and evaluation on the extensive literature on graphical perception and visualization effectiveness, summarized in the following section.

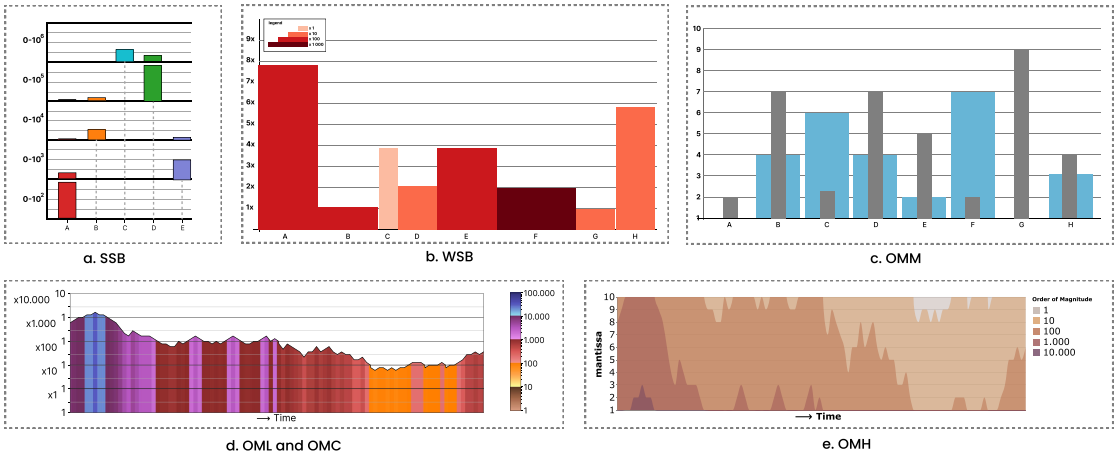

Figure 3: Mantissa-exponent visualizations: (a) Scale-Stacked Bar Charts (SSB) [29], (b) Width-Scale Bar Charts (WSB) [30], (c) Order of Magnitude Markers (OMM) [6], (d) Order of Magnitude Color and Line Chart (OML) [10], (e) Order of Magnitude Horizon Graph (OMH) [10].

Reference	Data Type	Mark Type	Range for Exponent	Channel for Exponent	Channel for Mantissa	Tasks Evaluation
SSB [29]	nominal	line	5 (10^0–10^4)	Row	Y position	Value, Sort, Difference, Ratio
WSB [30]	nominal	line	4 (10^0–10^3)	Length and Color Intensity	Y position	Value, Sort, Ratio, Trend
OMM [6]	nominal	line (2)	5 (10^1–10^5)	Y position and Color Hue	Y position and Color Hue	Value, Sort, Ratio, Trend
OMC [11]	timeseries	area	7 (10^{-8}–10^{-2})	Color Hue	Color Intensity	Value, Sort, Difference
OML [10]	timeseries	area	6 (10^0–10^5)	Y position and Color Hue	Y position and Color Intensity	Value, Sort, Difference, Trend
OMH [10]	timeseries	area	5 (10^0–10^4)	Color Intensity	Y position	Value, Sort, Difference, Trend

Table 1: Description of mantissa-exponent visualizations proposed in the literature.

2.3 Graphical Perception and Visualization Effectiveness

The concept of visualization effectiveness has evolved significantly since Bertin first introduced it [4]. Building on Bertin's foundation, Cleveland and McGill [18] conducted experiments to evaluate the ranking of visual channels for quantitative data, with a specific focus on the accuracy—how precisely a value can be identified [51] of sorting and comparison tasks. Mackinlay [42] expanded this ranking to include both nominal and ordered data, thus broadening our understanding of visual channel effectiveness. Several subsequent experiments [28, 35, 45] have empirically validated and refined the proposed effectiveness rankings of visual channels, revealing decoding factors, such as different tasks, that could influence them. Zeng and Battle [75], through a systematic review of the literature on graphical perception, pointed out that position encodings are the most accurate channels across all data types. The *accuracy* of other visual channels (e.g., area, length, or color intensity) varies, influenced by factors like *discriminability*—the ability to perceive distinct values for each attribute [51]—, and *separability*—the independence of channels from one another [51].

Discriminability is affected by several factors, including the number of objects [27], the type and size of a mark [41, 67], the background color [41], visual attention limitations [27], and interference effects with other channels [67]. Interference effects with other channels lead to separability issues. Separability allows for focused attention on one visual channel without interference from another [24]. Separable visual channels, such as position and color, show minimal cross-channel effects, causing no effectiveness issues when assigned to different attributes of a visualization. In contrast, when the encoding of one channel affects the perception of another, the channels are called *integral* [24]. Integral visual channel pairs include color hue and color intensity [15, 26], height and width [23, 37, 43, 68], shape and size [65], shape and color intensity [65], size and color intensity [67], and size and position [35]. The redundant combination of integral visual channels for the same attribute facilitates decoding. However, encoding two unrelated attributes using two integral channels can lead to interference, hindering identification and value decoding. [1].

Visualization researchers have leveraged this knowledge to develop quality metrics [3], heuristics [76], design guidelines [7], and recommendation systems [50], which are used to design visualizations and evaluate their effectiveness. Although some of these heuristics lay the groundwork for our exploration, they are not yet sufficient to determine whether a visualization that separates OMVs into mantissa and exponent is effective. In such visualizations, the perceptual task of magnitude estimation depends on the combination of two values—mantissa and exponent—rather than a single one. As a result, well-established low-level perceptual quality metrics for magnitude estimation may not fully apply in this context.

Potential interference between the visual encodings of mantissa and exponent could complicate the estimation process. Additionally, disregarding familiar mental models of numbers, such as the continuity between mantissa and exponent that mirrors verbal number expression, may hinder accurate comparisons.

This gap underscores the necessity for systematic exploration and evaluation of how mantissa and exponent are visually encoded and how these encodings are decoded perceptually and cognitively in various tasks. By identifying and documenting encoding and decoding challenges, our research aims to enrich the graphical perception literature by introducing guidelines that specifically support low-level perception and cognition tasks such as magnitude estimation and quantitative comparisons for visualizations that separate OMVs into mantissa and exponent.

3 DESIGN SPACE

In this section, we describe a design space for visualizing tabular data with one quantitative OMV, separated into mantissa and exponent. We follow a constraint-based approach [5], setting specific boundaries and refining the space by insights gained from the literature, related to: 1) the GoG [74] 2) studies on graphical perception [1, 54] and visualization effectiveness [4, 50, 75], and 3) the literature on perception [17, 39, 40] and visualization [6, 10, 30] of large value ranges.

Considering the combinatorial nature of all the choices a designer can make to visualize mantissa and exponent—such as choosing between different marks, visual channels, layouts, developing novel scales, employing redundant encoding, and adding design embellishments— describing, exploring, and evaluating the entire design space of all candidate solutions is infeasible. Therefore, we initially focused on a well-structured [64] subset of the design space, including visualizations with high data-to-ink ratio (no redundancy, single mark, no embellishments-with exception grid lines and tick marks), supported by current implementations of the GoG, such as Vega-Lite [61] and Plot [52]. This approach allowed us to observe the effectiveness of assignments for mantissa and exponent while minimizing interference from external design factors and to systematically generate a large number of visualizations.

We start by establishing the boundaries of our design space, proceed to describe its dimensions—how each visualization within this space is described—then we define the `EplusM` scale, and conclude by outlining the constraints implemented to refine the design space, with the goal of improving visualization effectiveness.

3.1 Boundaries

Our design space's boundaries are shaped by the objectives of our research and insights gained from the literature review. We concentrate on tabular datasets with a single OMV attribute, excluding types such as networks, fields, and geometries. Secondly, we focus on the Cartesian coordinate system, excluding polar or parallel layouts. Aiming to enhance the effectiveness of OMV visualizations in overview and design visualizations that are applicable to both printed and digital media, we limit our scope to static visualization designs and exclude interactive elements. Moreover, our design space does not entail data transformations; these are handled during dataset creation. Our analysis explicitly targets visualizations that employ a single mark, and consequently, we are removing the Order of Magnitude Markers [6] from our design space.

3.2 Dimensions

After setting the design space boundaries, we define the dimensions that specify each visualization. GoG implementations [52, 73] describe a visualization as the combination of data types, marks, visual channels, facets, scales, transforms, and projections. Considering the boundaries established, we describe each visualization design by a combination of the following elements: data, marks, and visual channels, considering faceting as part of visual channels [50, 61, 75]. Figure 1 shows the dimensions of our design space.

Data. In this context, "data" pertains to the type and range of attributes. We consider OMV as a specific type of quantitative data attribute that is separated into two quantitative attributes: **mantissa** and **exponent**, as defined in the background section. We explore visualizations that display the mantissa, the exponent, and an additional attribute type (we call *attribute-2*) among **nominal**, **ordinal**, **temporal**, and **quantitative**. For the exploration of the design space we created datasets with one-to-one relationship between the OMV and the attribute-2. Figure 1 displays the DATA (blue) part as rows on the top left of the table.

Marks. Marks are defined based on their spatial dimension requirements: **points** represent zero-dimensional marks, **lines** correspond to one-dimensional marks, and **areas** are considered two-dimensional marks [4, 51]. Within the framework of the GoG implemented by Plot, we represent points either with the Dot or with the Rule mark—when the visual channel of Length is chosen—, for lines, we use the Bar mark for nominal and ordinal data, the Rect mark for time, and last, for areas, we use the Area mark. Figure 1 displays the MARKS (green) part as rows on the left of the table.

Visual channels. Visual channels (VC) are ways to control the appearance of marks [51]. In our design space, we selected channels based on insights from the literature on visualization effectiveness [50, 75]. We focus on the following nine visual channels: X position (PosX), Y position (PosY), Row, Column (Col), Length, Area, Color Intensity (Intensity), Color Hue (Hue), and Shape. Length refers to both width and height, depending on the orientation of the line and the choice of the mark. Row and Col are channels derived from the faceting operation [50, 61, 75]. Figure 1 displays the VISUAL CHANNELS (red) as columns on the top of the table.

$$\text{VC} = \{\text{PosX}, \text{PosY}, \text{Row}, \text{Col}, \text{Length}, \text{Area}, \text{Intensity}, \text{Hue}, \text{Shape}\}$$

Description of Visualizations. Considering the dimensions above, we describe a visualization that belongs to our design space as a list of four *assignments* forming a *visualization configuration*:
Mark, Exponent $\mapsto$ VC, Mantissa $\mapsto$ VC, Attribute-2 $\mapsto$ VC.

Figure 1 displays the combination on the bottom of the generated visualization as:

Figure 4: Point, Exp. $\mapsto$ Row and Intensity, Mant. $\mapsto$ PosY, Nominal $\mapsto$ PosX

Scales. According to the GoG, a visualization scale is a function that transforms attribute values into visual channel values, such as position, length, or color intensity. In our design space, scales are handled by Plot and are determined automatically by the attribute type. Band scales are utilized for nominal attributes, ordinal scales for ordinal attributes, and linear scales for quantitative attributes. To simplify interpretation and eliminate the need for calculations, we implement a tick transformation representing the exponent value as 10^e. For instance, if the exponent is 3, the tick label displays 1,000 (10^3). Given that scales are selected based on the data attribute type, they are not discussed within our design space, with the exception of `EplusM` (see below).

3.3 `EplusM` Scale

As mentioned in the background, studies in cognitive psychology [40] suggest that people perceive OMVs following a segmented linear model, with magnitudes distributed uniformly as in logarithmic scales and numbers within these magnitudes being linearly interpolated. We believe the scale used by Braun et al. [10], and inspired by the positional encoding of SSB [29], aligns closely with this segmented linear model. However, while Braun et al. [10] encode exponent and mantissa continuously on a single axis, they still display the two components in y-axis separately (see Figure 3d), requiring users to multiply them to derive the corresponding value.

We integrate the mantissa and exponent into a single positional scale, termed `EplusM` (exponent plus mantissa) and refine its definition to facilitate experimental evaluation and encourage broader adoption in OMV visualizations. `EplusM` is a piecewise linear scale that maps a value (v) to visual elements using the formula: $s(v) = e(v) + (m(v) - 1)/9$, where $e(v)$ is the exponent and $m(v) \in [1, 10)$ is the mantissa. This scale uses the exponents as primary units and fills the $[0, 1)$ interval between consecutive exponents, by rescaling the mantissa to $(m-1)/9$. When applied to positional encoding, the axis units (major ticks) represent the values of the exponents, and the subunits (minor ticks) correspond to the mantissa values. A tick transform is necessary to show the original value v corresponding to each $s(v)$. We recommend displaying the midpoint between two exponents to highlight the linearity of the scale and illustrate the segmented linear model. The supplemental material includes an example of `EplusM` implementation. In our design space, this scale is described by assigning both the exponent and mantissa to the same positional channel (e.g., Exponent $\mapsto$ PosY, Mantissa $\mapsto$ PosY).

3.4 Constraints

To ensure the effectiveness and expressiveness of the visualizations in our design space, we apply constraints related to the boundaries we have set, the expressiveness of marks and visual channels, and the interference of different channels. We consider a visualization *viable* only if all constraints are satisfied. Our approach is inspired by the literature on recommendation systems for visualization [50], where integrity constraints are used to encode expressiveness and restrictions to the dimensions of a visualization specification.

Constraints based on layout. As we limit our design space to the Cartesian system, we include only combinations that contain at least one position channel (PosX or PosY).

Constraints for the expressiveness of channels. To ensure the expressiveness of each assignment [51], we use Hue and Shape only for nominal data, Intensity for ordinal and quantitative data, and Length and Area only for quantitative data. For the Row and Col channels, we assign only the exponent, as mantissa is a quantitative attribute with continuous range, and attribute-2 has a one-to-one relationship with the OMV, leading to visualizations similar to those using a positional channel. An exception to this constraint is our decision to retain visualizations that include both row and column for the exponent and attribute-2 as a nominal attribute, as they result in novel visualizations.

Constraints based on mark type. For each type of mark, we impose constraints in accordance with its defined dimensions. The Shape and Area channels can be used only for point marks, and the Length channel is not feasible for area marks. We also apply constraints in relation to the dataset. We use area marks only for temporal data, and we exclude lines for quantitative data.

Constraints based on interference with identity channels. Channels like Hue and Shape serve identification purposes. Combining Hue (e.g., red) with Intensity within the same mark generates a new color (e.g., pink). Similarly, merging Shape with Length can alter the shape (e.g., modifying a circle into an ellipse). These combinations modify the identity of the primary channel, and thus, we do not consider combining Hue with Intensity and Shape with Length.

4 GENERATION OF VISUALIZATIONS

We developed a tool that enabled us to systematically generate and explore all possible assignments of visual channels to mantissa and exponent across marks and data attribute types. Figure 5 shows examples of the generated visualizations.

4.1 Order of Magnitudes Tool

To facilitate the exploration of the OMVs design space, we implemented a web-based open-source tool, OMVis, written in JavaScript. Our tool includes two main components: 1) a menu for the interactive exploration of all possible combinations and 2) the resulting visualization. Within the menu, users can upload or select a dataset from our database. They then choose an OMV to separate into its mantissa and exponent parts, followed by the selection of additional attributes (nominal, ordinal, time, or quantitative). Subsequently, users decide on a mark (point, line, or area) and assign visual channels to the mantissa, exponent, and selected attributes. We apply the constraints by disabling the corresponding choices; for example, quantitative visual channels such as Length are not applicable to nominal attributes. A visualization is generated based on these selections. We use Plot to create the visualizations. The user can export the produced visualization either as an image or as JavaScript code. More details are provided in the supplemental material.

4.2 Datasets

We used the "Trending YouTube Video Statistics" dataset from Kaggle [33], which comprises daily metrics (views, likes, comments) for the top trending YouTube videos across various categories. We selected this dataset because it contains OMVs (e.g., # of views)

Figure 5: Two examples of visualizations per dataset: (a) Nominal, (b) Quantitative, (c) Temporal, and (d) Ordinal. All the visualizations were generated using our OMVis Tool. At the top, visualizations are described according to the design space dimensions (see Figure 4).

and includes diverse attributes such as nominal (e.g., video categories), ordinal (e.g., 7-point Likert scale sentiments), temporal (e.g., days), and quantitative (likes, comments). For the design space exploration, we created four smaller datasets (maximum cardinality = 18), each featuring views (OMV) and another attribute: nominal, ordinal, temporal, or quantitative. The exponent range is from five to eight. The datasets used in our exploration are available in the supplemental material.

4.3 Visualizations Generation Process

We systematically generated all visualizations by first calculating every combination of the nine visual channels ($n = 9$) in our design space with our data attributes ($k = 3$) {*exponent*, *mantissa*, *attribute-2*}, where *attribute-2* represents a nominal, ordinal, temporal, or quantitative attribute. For our exploration, we considered only one-to-one assignments, where one data attribute is encoded with one visual channel. This constitutes a permutation problem since the order of assignment matters (i.e., which visual channel is assigned to which data attribute). The formula for permutations of n items taken k at a time is given by: $P(n,k) = \frac{n!}{(n-k)!}$ so, in our case, $P(9,3) = 504$ visualizations.

An exception to the one-to-one rule involves the visual channels of PosX and PosY, where both the mantissa and exponent can be assigned to the same visual channel, resulting in the scale denoted as `EplusM`. The total number of combinations for the `EplusM` scale for both PosX and PosY is $2 \times P(8,1) = 16$ visualizations.

This method was uniformly applied across all mark types {*point*, *line*, *area*} and data attribute types for attribute-2 {*nominal*, *ordinal*, *time*, *quantitative*}, resulting in the total number of combinations before imposing any constraints being equal to:

$$3 \times 4 \times (504 + 16) = 6,240$$

The result after the application of constraints mentioned in the previous section was 336 viable combinations. We used our tool to generate all the viable visualizations and exported them as images. For details about the creation of the combinations and the applications of constraints, see the program `combinations.js` in the supplemental material.

5 QUALITATIVE EVALUATION

We qualitatively evaluated all generated visualizations using a systematic, inspection-based method inspired by the heuristic walkthrough technique [63]. This two-pass evaluation approach guides evaluators with a set of heuristics, a prioritized list of user tasks, and "thought-focusing" questions [63]. During each pass, evaluators document problems, assign individual ratings, and later compare and discuss the identified issues.

Building on this technique, we conducted two evaluation sessions. The first, called **Encoding Walkthrough**, was a group evaluation where we displayed all visualizations on a wall-sized screen and documented issues related to the encoding of OMVs, guided by visualization-specific questions and heuristics. The second, called **Decoding Walkthrough**, involved individual evaluations where

each evaluator completed specific tasks, rated the visualizations, and noted any problems affecting the effectiveness of decoding the visualized OMVs.

Finally, we entered a collaborative phase to discuss and compare our ratings and observations. At the end of this process, we classified each visualization into four categories based on the identified problems and presented the insights gained as a set of design guidelines for visualizing mantissa and exponent.

5.1 Evaluators

The group of evaluators (henceforth referred to as "we") who conducted the qualitative evaluation consisted of four experts in data visualization and/or human-computer interaction, all of whom are co-authors of this paper: one doctoral student, one practitioner and two faculty members. Each evaluator has a range of 5-30 years (mean=14 years) of research and practical experience in data visualization and/or human-computer interaction. Our coding methodology adheres to the tradition of expert-based evaluation of user interfaces, such as heuristic evaluation and cognitive walkthrough, where it has been shown that a small number of experts can reliably identify most problems [44, 72].

5.2 Encoding Walkthrough

The focus of this session was to identify problems relating to the visual encoding of OMVs. This evaluation was guided by visualization heuristics related to the effectiveness of visual channels [51, 69, 76] with a focus on aspects that could impact magnitude estimation (see Section 2.3). We displayed all the visualizations on a wall-sized screen (6m wide and 2m high) and conducted a "walkthrough" of each. For every visualization, we aimed to answer the following heuristics-inspired thought-focusing questions:

- **Accuracy:** How precisely can we tell the difference between the different values of each data attribute?
- **Discriminability:** How many unique values can we perceive for each attribute?
- **Separability:** Do the channels interfere with each other?
- **Grouping:** Can we identify values belonging to the same exponent?
- **Continuity:** Do the channels used for the mantissa and exponent convey the numerical relationship between the two attributes?

We labeled each visualization based on whether or not we observed issues related to the previously defined questions regarding the mantissa, exponent, or the attribute-2, while also describing, the specific identified problems. For example, if there was a problem related to the separability between channels, we labeled separability with "no", providing more details about the problem's origin (e.g., interference effect between Intensity used for attribute-2 and Area for exponent). Additionally, we evaluated the suitability of the displayed encodings for visualizing OMVs, responding with "yes," "no," or "maybe." Using the Miro tool [49], we then grouped the 336 visualizations according to their suitability labels (yes, maybe, no) and identified patterns related to the effectiveness of different visual encodings, taking into account the problems observed. Table 2 shows a summary of the encoding-identified problems.

We found that several identified issues were caused by the high cardinality (18) of the chosen datasets, impacting the readability of visualizations and hiding potential problems. As a result, for the second stage of our evaluation process, we reduced the cardinality to 7—small enough to ensure discriminability across all channels—by selecting representative values from each order of magnitude. Additionally, we observed that all symmetrical visualizations shared similar strengths and weaknesses. To streamline the evaluation, we removed the duplicated symmetrical visualizations, reducing the total number of visualizations from 336 to 168. We prioritized symmetrical combinations involving both position and faceting, opting to retain PosY and Row over PosX and Col [19, 35].

5.3 Decoding Walkthrough

The focus of this session was to identify issues affecting the decoding of visualized OMVs, specifically in relation to three tasks: value retrieval, difference estimation, and ratio estimation. Based on the literature on visualizing OMVs [10, 29, 30], we considered three conditions for difference and ratio estimation based on the relationship between the compared values: within the same, neighboring, or distant orders of magnitude (i.e., the exponent of the values differ by at least two).

To conduct the evaluation, we developed an online tool. We assigned 84 visualizations to each evaluator, and therefore each visualization was evaluated by a pair of evaluators. Each evaluator performed the defined tasks and provided ratings on perceived accuracy for both the exponent and mantissa, the effort required to complete the task, the confidence in the provided answers, and the overall suitability of the visualization for OMVs. Additionally, evaluators could leave comments on any task-related issues they observed. Following the individual coding phase, we held a collaborative session to discuss and review all ratings. Inspired by collaborative coding methods in thematic analysis [12, 57], this collaborative session aimed to compile similar identified problems and reach a common agreement for all the visualizations.

5.4 Results

The qualitative evaluation yielded a set of observed problems and labels for each visualization regarding both encoding and decoding (see Table 2). Based on these insights, we categorized the 168 visualizations into four groups:

Effective assignments (25/168): Visualizations labeled as suitable for OMVs had no significant encoding issues and only minor decoding caveats that could be resolved with fine-tuning or design enhancements. All visualizations in this category utilized a positional channel (PosY or Row) for the exponent, paired with a highly accurate channel for the mantissa (PosY or Length) aligned in the same direction, ensuring visual continuity. Figure 6 shows three examples of these visualizations.

Value retrieval, but no comparisons (7/168): Visualizations effective for value estimation but requiring extra effort for quantitative comparisons. While they use highly accurate and discriminable channels for both mantissa and exponent, the lack of continuity forces users to multiply the two to derive the value (Figure 7a).

Figure 6: Sample of "Effective assignments" combinations (3/25) . At the top, visualizations are described according to the design space (see Figure 4), while at the bottom are abbreviations for the guidelines (see Section 5.5) each visualization follows (✓). This sample of visualizations use positional channels for both the mantissa and exponent: the first two demonstrate the EplusM scale on the y-axis, and the third provides an example of Facet (the exponent is encoded with Row, and the mantissa with PosY).

(a) "Value retrieval, but no comparisons"

(b) "Magnitude overview, but no detail"

(c) "Impractical for OMVs"

Figure 7: Examples of combinations that violate guidelines, labeled based on the group to which they belong. At the top, visualizations are described according to the design space (see Figure 4), while at the bottom are abbreviations for the guidelines (see Section 5.5) each visualization follows (✓) or violates (✗).

Magnitude overview, but no detail (55/168): Visualizations that effectively communicate the magnitude of a value by using a highly accurate channel for the exponent, but the encoding of the mantissa with a low-discriminability channel (e.g., Intensity or Area) requires moderate effort to perform tasks and leads to low confidence in the results (Figure 7b).

Impractical for OMVs (81/168): Visualizations that are unsuitable for visualizing OMVs. This group includes those that lack a highly accurate channel for the exponent or exhibit interference between channels, making them unsuitable for tasks including OMVs (Figure 7c).

Through our analysis, we identified common patterns that we consider necessary for the effective visualization of OMVs. We present these patterns as design guidelines, which are detailed in the following section. The 168 images used in our evaluation are available in the supplemental material.

5.5 Design Guidelines

Based on the results of the evaluations, we established four guidelines for effective visualizations. The first, **Accuracy for Magnitude (AcM)**, requires encoding the exponent with a highly accurate and discriminable channel to avoid significant errors in value estimation and comparisons, which is crucial for visualizing OMVs. The second, **Detail Inside Magnitudes (DeM)**, stresses the need for the mantissa to also use an accurate and discriminable channel, enabling precise comparisons and estimations of differences and ratios within a single magnitude. Failure to follow this guideline reduces accuracy. **Continuity Between Magnitudes (CoM)** ensures a smooth transition between exponents by using a mantissa channel that mirrors the numerical relationship between the exponent and mantissa. This ensures a logical progression across values and aids in comparing neighboring exponents. Lastly, **Parsimony in Channels (PaC)** recommends using the fewest possible channels to represent both attributes. This minimizes channel interference,

Encoding Problems	Decoding Problems
Low accuracy/ discriminability for exponent	Increased effort to perform the tasks, potentially leading to an order of magnitude error.
Low accuracy/ discriminability for mantissa	Decreased confidence for value estimation and increased effort for quantitative comparisons.
No continuity between mantissa and exponent	Increased effort for quantitative comparisons; the user needs to retrieve and multiply mantissa and exponent to derive the value.
Low accuracy/ discriminability for the other attribute	Reduced readability, increased effort to perform the tasks.
Interference effects	Reduced readability, increased effort to derive the value.
No grouping for exponent	Increased effort for quantitative comparisons.
Encoding constraints based on data cardinality	Reduced readability. Constraints in visualization size and visual channels that can be used for the other attribute.
Encoding constraints based on mark	Constraints in visual channels that can be used due to interference effects with the mark. Potential need for redundant encoding to increase effectiveness.
Encoding constraints based on data characteristics.	Constraints in visual channels that can be used. Need for design embellishments to address and communicate potential issues to users.

Table 2: Encoding and corresponding decoding problems that we observed during the qualitative evaluation.

and allows for the effective integration of additional data attributes into the visualization.

Overall, two encodings for mantissa and exponent that adhere to the mentioned guidelines and appear to result in effective visualizations for OMVs are the **`EplusM`** (see Figure 8d) scale, which encodes both the mantissa and exponent with the same positional channel and the **Facet** (see Figure 8e), in which the exponent is encoded with row and the mantissa with Y position. To validate that our guidelines can improve the effectiveness of visualizations that separate OMVs into mantissa and exponent, we conducted an empirical study described in the following section.

6 CONTROLLED EXPERIMENT

To test whether adhering to our proposed guidelines leads to more effective visualizations, we assessed the effectiveness of EplusM and Facet in a crowdsourcing experiment using Prolific [53]. This study compared `EplusM` and Facet with two baseline conditions: linear and logarithmic scales, as well as a well-established visualization from the literature, the SSB [29]. We selected SSB because it is the first and most widely adopted visualization to split mantissa and exponent for OMVs, and it fulfills most of our proposed guidelines.

We expect our proposed solutions to outperform SSB for two primary reasons. While SSB uses positional channels for both mantissa and exponent, adhering to two of our proposed guidelines, we argue that the repetition of values in each stacked scale violates the guideline of maintaining continuity between magnitudes. Additionally, the use of color to represent categories breaks the guideline of parsimony in the use of visual channels. As SSB already outperformed Lin and Log [29], we expect our designs to outperform them as well.

We formulated the following hypotheses:

- $\mathbf{H_{err}}$:*Our designs exhibit similar or lower error rates for tasks on OMV visualizations compared to Lin, Log and SSB.*
- $\mathbf{H_{tim}}$:*Our designs exhibit similar or shorter response times for tasks on OMV visualizations compared to Log and SSB.*
- $\mathbf{H_{con}}$:*Our designs exhibit similar or increased confidence for tasks on OMV visualizations compared to Lin, Log and SSB.*

6.1 Data Generation

In our empirical evaluation, we focused on the example of a national budget, a common case of a dataset with a wide range of values spanning multiple orders of magnitude. Specifically, we used the French National Budget, which includes programs with OMVs ranging from 10 thousand ($e = 4$) to 100 billion Euros ($e = 11$), covering seven orders of magnitude. The use of such data extends the range of orders of magnitudes previously studied (4 for [30] and 5 for [6, 10, 29]) and allows the use of real numbers to represent mantissa—unlike previous ones who used only integers.

To test the effectiveness of the visualization designs under these conditions, we generated synthetic datasets that reflect this extensive value range. Each dataset included 14 categories, with two categories assigned to each exponent level, similar to previous studies [30]. This design allowed for comparisons both within the same exponent and across neighboring exponents. The mantissa for these categories was randomly generated within the range $[1, 10)$, and each category was labeled with a unique letter from the alphabet. In total, we generated 500 statistically equivalent datasets. Every participant performed the trials with a different dataset, to ensure the results are not overly influenced by potential artifacts in the characteristics of one generated dataset per task. While we do not claim ecological validity with these synthetic datasets, they provide controlled conditions that highlight potential challenges and issues related to OMVs in different visualizations.

6.2 Experiment Stimuli

Given that our dataset focuses on a national budget, we required a visualization that effectively supports both nominal data (the budgeting categories) and OMVs (the budget amounts for each category). We opted for a line mark, as bar chart designs are among the most common visualization methods for this type of data [14]. For the visual channels, we chose the positional encodings of `EplusM` and Facet, as these encodings exhibit the least number of problems during our qualitative evaluation and reflect well our proposed guidelines. We implemented the visualizations using d3.js [9] and ensured consistent size dimensions with Sass [60] and React [55].

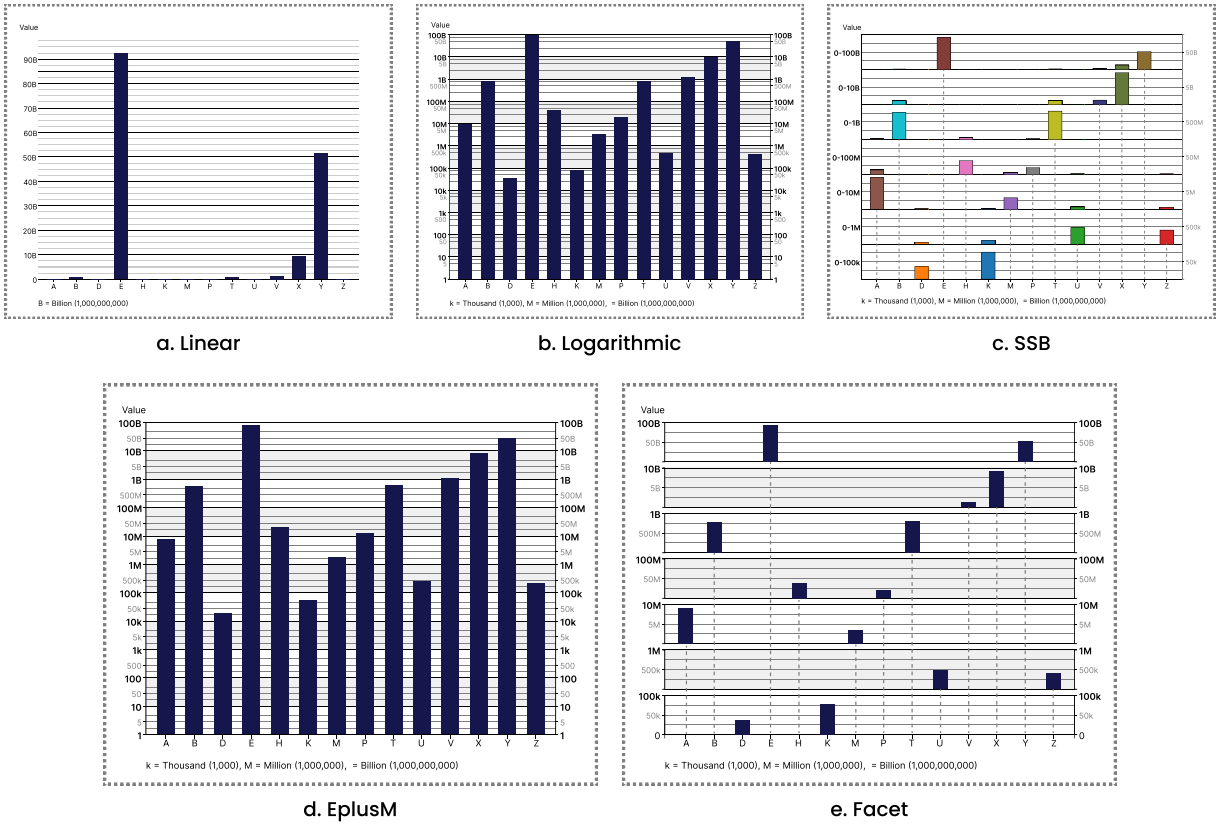

Figure 8: Visualization designs used for the crowdsourcing evaluation: (a) Linear bar chart, (b) Logarithmic bar chart, (c) Scale-Stacked Bar Charts (SSB) [29], (d) EplusM bar chart, (e) Facet bar chart. All visualizations display the same dataset.

We examined five visualization designs: the Linear Bar Chart, Logarithmic Bar Chart, Stacked-Scale Bar Chart (SSB), `EplusM` Bar Chart, and Facet Bar Chart (Figure 8). We refined the `EplusM` and Facet Bar Charts using state-of-the-art design guidelines [51], while the SSB design was faithfully replicated in its original form. To ensure comparability across all visualizations, we standardized key design elements such as the font, font size, grid lines, and overall styling. For instance, we used thick black horizontal lines to separate scales by exponent and thin grey lines to subdivide each scale into four equal parts, corresponding to mantissa values of 2.5, 5, and 7.5, consistent with the SSB design. In addition, we added a single tick at the mantissa value of 5 in all designs, alongside the existing ticks used to differentiate between exponents. While we maintained a uniform color scheme across all visualization designs, we preserved the original color encoding in the SSB design. We believe that the redundant encoding for category in SSB does not impact user performance, as it is not related to OMVs retrieval or comparison. These consistent design choices ensured a fair assessment of each visualization design's impact on participant performance, avoiding confounding factors.

6.3 Tasks

We focused on the following low-level perception and cognition tasks for magnitude estimation and quantitative comparison:

Value retrieval. The objective is for participants to estimate a value as accurately as possible. This task forms the first block of our experiment, consisting of four trials.

Difference estimation. Participants aim to estimate the difference between two values as accurately as possible. This task constitutes the second block of our experiment and builds on the approach by Braun et al. [10]. We structured the task into three subtasks based on the relationship between the two values: belonging to the *same* order of magnitude, *neighboring* orders of magnitude, or *distant* orders of magnitude.

Ratio estimation. Participants aim to estimate the ratio between two values as accurately as possible. This task constitutes the third block of our experiment, and similarly to the difference estimation task, we structured the task into three subtasks based on the relationship between the two values: *same*, *neighboring*, and *distant*.

The three blocks were presented in the same order for every participant, from the easiest to the most difficult: 1) Value retrieval, 2) Difference, and 3) Ratio. Within the last two blocks, the three tasks also followed a fixed order from the easiest to the most difficult: 1) *same* order of magnitude, 2) *neighboring* orders of magnitude, and 3) *distant* orders of magnitude. Each of the seven tasks included four trials.

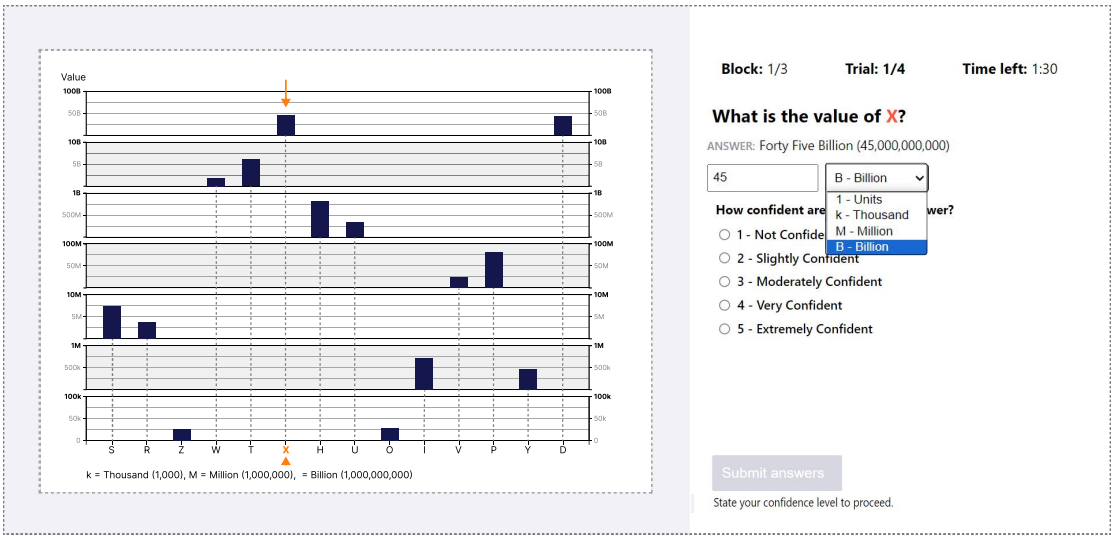
Value
S R Z W T X H U O I V P Y D
k = Thousand (1,000), M = Million (1,000,000), = Billion (1,000,000,000)
Block: 1/3
Trial: 1/4
Time left: 1:30
What is the value of X?
ANSWER: Forty Five Billion (45,000,000,000)
45
B - Billion
1 - Units
k - Thousand
M - Million
B - Billion
How confident are
wer?
1 - Not Confide
2 - Slightly Confident
3 - Moderately Confident
4 - Very Confident
5 - Extremely Confident
Submit answers
State your confidence level to proceed.

Figure 9: Interface of the tool we developed to conduct the crowdsourcing experiment. Example of Value Retrieval task.

To help participants identify the selected categories, we highlighted them in orange and used arrows positioned above the corresponding bar and below the letter on the x-axis as additional visual cues. The answer input fields align with the format used in the visualization, featuring a numeric input field for values between 1 and 999, along with a dropdown menu to select the appropriate unit (thousands (k), millions (M), or billions (B)). We displayed the participant's response above the input fields in both words and numbers, reinforcing their input. Figure 9 shows an example of the interface used for the value retrieval task.

6.4 Measures

We evaluated and compared the visualization designs using three measures: accuracy, response time, and subjective confidence. For accuracy, previous studies involving tasks with precisely determinable responses have used two error measurement methods:

Absolute Relative Error = $\left|1 - \frac{response}{correct}\right|$. This error metric, used by Hlawatsch et al. [29] and Braun et al. [10], measures the relative deviation from the correct value. It has some advantages for measuring accuracy with OMVs as it takes into account the characteristics of logarithmic scaled data. For example, $\left|1 - \frac{10}{100}\right|$ and $\left|1 - \frac{1000}{10000}\right|$ both yield an error of 0.9, despite the absolute errors (90 and 9000) being different. However, this method is asymmetrical. It produces a lower error when the participant underestimates the correct value, and a higher error when they overestimate. For example, if the correct value is 100 and the response is 10, the error is 0.9, but if the correct value is 10 and the response is 100, the error jumps to 9.

Log Relative Error = $\log_{10}\left(\frac{response}{correct}\right)$. This error metric (hereafter abbreviated in $e_{\log}$), used by Höhn et al. [30], calculates the logarithmic relative difference between the correct and response values. It is symmetric [71], meaning that the same order-of-magnitude error yields the same value regardless of whether the participant overestimates or underestimates. Tofallis [70] further supports that it is the most accurate measure of relative differences for heteroscedastic data, which is particularly relevant for OMVs, where the response and the correct value can differ significantly in orders of magnitude.

Given these considerations, we used $e_{\log}$ to measure accuracy, comparing participants' responses to the "visually correct value," defined as the highest level of detail the visualization conveys. Based on our data range and visualization dimensions, the visually correct value was rounded to two digits after mantissa. For analysis, we took the absolute Log Relative Error since the direction of the error was not of interest. The $e_{\log}$ is 0 when the response matches the correct value, and 1 when the error is off by one exponent. In value retrieval tasks, a Log Relative Error between 0 and 1 indicates an error in estimating the mantissa. Since the logarithm of 0 is undefined, we replaced any participant responses of 0 (equivalent to NaN or null) with 1 for calculation purposes.

6.5 Experimental Design

We conducted the study online using the Prolific.com [53] service. We employed a between-subjects design for the different visualizations and a within-subjects design for the different tasks [16]. This approach reduced the number of trials per participant, allowing us to distribute tasks across various data conditions and enabling a more in-depth analysis of quantitative comparisons for values spanning multiple orders of magnitude. To conduct the experiment, we developed a custom tool.

6.5.1 Pilot. We conducted two pilot studies before the main experiment. In the first pilot, we recruited 30 participants through personal contacts to gather feedback and estimate the approximate duration of the experiment. Based on their input, we made several adjustments to both the instructions and the interface design. For the second pilot, we involved 22 users from Prolific to identify any technical issues that needed to be resolved before the main study. This pilot revealed responsiveness issues with specific screen dimensions, which we promptly addressed.

6.5.2 Participants and exclusion criteria. We conducted a power analysis to determine the number of participants required for our study. Using G*Power [22], we assumed an effect size of 0.25, an alpha level (α) of 0.05, and a power of 0.81. The analysis was based on 5 groups and 4 measurements per task (trials), with a correlation of 0.5 between measurements. Based on these parameters, the total sample size needed is 130, which equates to 26 participants per visualization. We recruited participants in an IRB-approved study on Prolific.co who are between the ages of 18 to 70, have a normal or corrected-to-normal vision, speak fluent English, had an approval rate 99%, and had completed more than 10 previous studies. We excluded participants who failed attention check questions, did not complete the study, had three consecutive missing answers, or achieved an average performance rating of fewer than two stars (indicating a lack of understanding of the experiment). We also excluded one participant, who wrote in the feedback section that he encountered technical problems during the experiment. In total, data were collected from 150 participants. After excluding 10 participants for failing attention checks, 3 for missing trials, 4 for low performance, and 1 who faced technical issues, 132 participants remained for analysis. Successful participants received compensation at an average rate of £9.97 per hour, with a median completion time of 31 minutes. The participants' mean age was 25 years, with 41 women, 55 men, and 36 who did not consent to share this information.

6.5.3 Crowdsourcing Quality Control. To ensure the integrity of our online experiment, we established strict hardware and software requirements for participants, aiming to reduce as much as possible the influence of confounding variables. Specifically, we required devices with a screen resolution of at least 1440 pixels in width and 850 pixels in height. We calculated visualization sizes in centimeters, accounting for each device's pixel ratio and zoom level. To maintain consistency in the visual presentation, we restricted participants from adjusting their browser's zoom level. If the zoom level deviated from 100%, a full-screen message prompted them to reset it before continuing the experiment. We included a comprehension test after the instructions, allowing participants two attempts to answer correctly, and embedded three attention checks throughout the experiment. Additionally, we tracked how many times participants changed or minimized their window as an extra measure of attention. Finally, we incorporated a star rating for performance as a gamification element to enhance engagement, and we awarded a bonus payment to the top 5% of performers in each condition. Potential uncontrollable external influences such as distracting environmental noise was averaged out due to the large number of participants.

6.5.4 Procedure. The entire study lasted approximately 30 minutes and was decomposed into 5 phases:

Introduction and consent. Each participant began by reading an overview of the study's goals, requirements, and constraints. They then read the consent form and decided whether to participate in the experiment.

Subjective Numeracy Test. Each participant assessed their own ability with various mathematical tasks using the Short Subjective Numeracy Scale [46, 47]. Researchers demonstrated this scale is a highly accurate proxy to measure numeracy. We added two additional Likert scale questions regarding familiarity with logarithmic scales, as well as an attention check [8].

Instructions. Each participant saw detailed instructions about how to read the specific visualization design assigned to them and what they had to do during the experiment. At the end of the instructions, there were two comprehension check questions to ensure the understanding of the instructions. They had two chances to answer correctly; if they failed, the study was stopped. The median time spent in this section was 8.28 minutes.

Trials. Participants had one and a half minutes to answer each question and indicate their confidence level for their answer. After this time, the system automatically advanced to the next trial. To prevent participants from rushing through the trials, a five-second minimum viewing time was enforced before they could proceed, ensuring they had sufficient time to engage with the visualizations. After submitting their answers, participants received a performance rating in the form of stars, offering feedback on their accuracy and increasing engagement. Each participant received a unique dataset, with different visualized data for each trial. At the end of each block, participants saw a screen displaying a summary of their performance using the same stars system and introducing the next task. These screens also included attention check questions. The median time spent on this section was 16.42 minutes.

Feedback. At the end of the experiment, participants were asked to rate their overall experience using a 5-point Likert scale, ranging from "Very Negative" to "Extremely Positive." Additionally, they were invited to provide written feedback on the effectiveness of the visualizations for the various tasks, any problems they encountered, and suggestions for improving the visualization designs.

The tool used for the evaluation and the instructions for each visualization are included in the supplemental material.

6.6 Results

The final number of participants included in our study after exclusion criteria was 26 for Lin, 27 for Log, 26 for SSB, 26 for `EplusM`, and 27 for Facet. Since all included participants completed every task, our analysis is based on 132 observations, corresponding to the total number of participants. There is a weak correlation (Spearman Correlation: -0.19, $p<0.05$) or no correlation between the experiment measurements and participants' subjective numeracy or familiarity with logarithmic scales. Therefore, we do not consider these aspects in our analysis. In the following paragraphs, we present the results of our experiment per measurement. We follow current APA recommendations [2] and report our results using estimation techniques with effect sizes and confidence intervals (CIs), and thus avoid words such as "significant" in the discussion of our results [20]. We used bootstrapping to compute all the CIs. The experimental data, the analyses from this experiment, and the tables with the results are available in the supplemental material.

6.6.1 Accuracy. For accuracy, we used the $e_{\log}$ measurement described in Section 6.4. For each participant, we aggregated their four trials using the median error as it offers a robust measure of central tendency, reducing the influence of extreme values caused

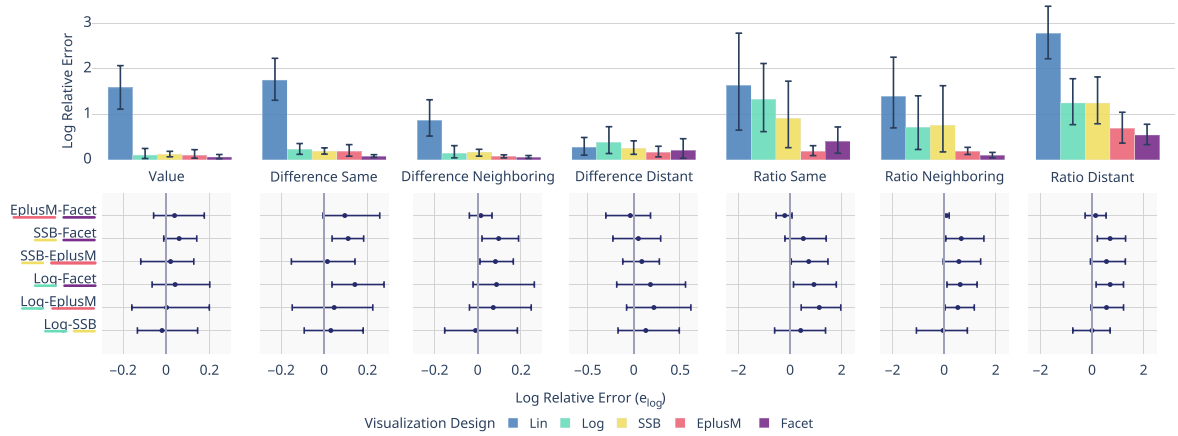

Figure 10: The bar chart shows the Log Relative Error ($e_{\log}$) per visualization design, grouped by task, with error bars representing the 95% CIs. For each task, we display the pairwise differences in $e_{\log}$ between Log, SSB, `EplusM`, and Facet with 95% CIs.

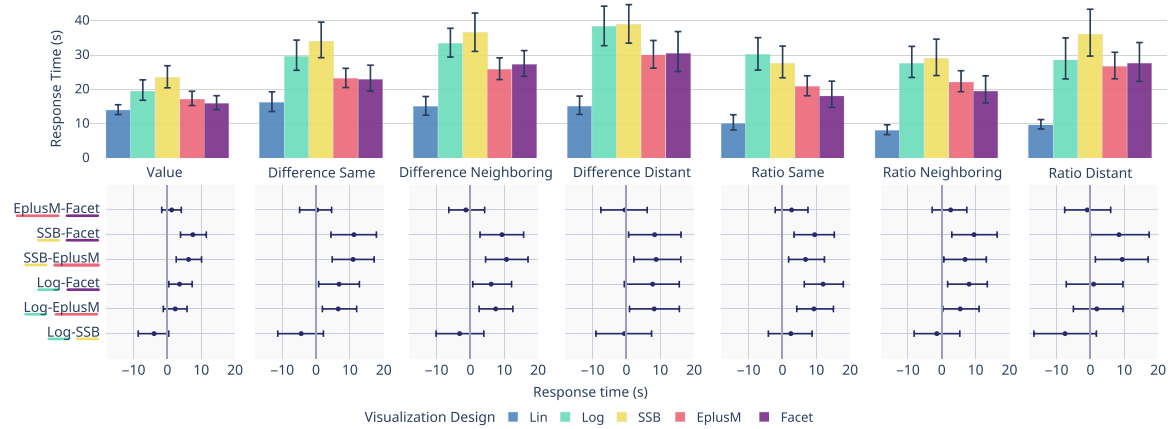

Figure 11: The bar chart shows the Response Time in seconds per visualization design, grouped by task, with error bars representing 95% CIs. For each task, we display the pairwise differences between Log, SSB, `EplusM`, and Facet with 95% CIs.

by large errors. Following Tofallis [70], we analyzed the $e_{\log}$ for each visualization and task using the geometric mean. The results (cf. Figure 10) indicate that Lin is the least accurate for most tasks, except for estimating differences between distant magnitudes. Consequently, and to maintain the readability of the charts, we chose not to perform pairwise comparisons for Lin.

Our results suggest that accuracy varies across tasks, with stronger effects observed for quantitative comparisons, where Facet and `EplusM` seem to exhibit lower error rates (H_{err}). We present a summary of key observations per task below:

- **Value retrieval:** Accuracy appears similar across Log, SSB, `EplusM`, and Facet.
- **Difference estimation:** For differences within the same magnitude, Facet demonstrates the highest accuracy, outperforming both SSB and Log. Similarly, for differences between neighboring magnitudes, Facet and `EplusM` outperform SSB and Log, with moderate evidence supporting this trend. However, for differences between distant magnitudes, the results are largely inconclusive.
- **Ratio estimation:** Evidence suggests that `EplusM` and Facet outperform Log and SSB across all magnitude conditions, with greater effects observed between `EplusM` and Log in the same magnitude condition.

6.6.2 Response Time. For each participant, we computed the geometric mean response time across their four trials. We selected the geometric mean because it effectively handles the skewed data typically seen in response times, reducing the influence of extreme values. Figure 11 presents the results. The results show that Lin has the fastest performance across all tasks. However, due to its consistently low accuracy, we chose not to perform pairwise comparisons for Lin.

Our results indicate that Facet and EplusM appear to be faster than SSB and Log (H_{tim}) across tasks, with a stronger effect observed for quantitative comparisons between values within the same or neighboring magnitudes. When comparing SSB and Log, the direction of the effect is uncertain but likely favors logarithmic scales, except in the case of ratio estimation for values within the same magnitude.

Figure 12: The bar chart on top shows the Confidence per visualization design, grouped by task, with error bars showing 95% CIs. For each task, the stacked bar charts on the bottom show the distribution of confidence for all trials, per design.

6.6.3 Confidence. For the visualization and analysis of Likert-scale responses, we follow the recommendations of South et al. [66]. Figure 12 (bottom) provides an overview of the Likert-scale answer distributions for confidence across all trials, tasks, and designs. In analyzing the results, we treat the responses as interval data, assuming equal distances between points. This intervalist approach is consistent with the analysis used for objective measures. We aggregate confidence ratings per participant by calculating the mean confidence level for each task. We then compute the overall mean and 95% confidence intervals for all visualizations per task, as shown in Figure 12 (top). This confidence rating serves as an indicator of the subjective effectiveness of the visualizations in supporting task performance, complementing the objective measures.

Our results suggest that using Facet and `EplusM` may lead to increased user confidence across tasks ($\mathbf{H_{con}}$). Log and SSB perform similarly across tasks with exception the estimating ratios within the same magnitude, where SSB appears to have slightly increased confidence. Log and SSB perform similarly, except for estimating ratios within the same magnitude, where SSB shows slightly higher confidence. Lin demonstrates the lowest confidence across tasks, except for estimating differences between distant magnitudes, where it seems to outperform both Log and SSB.

6.7 Result Discussion

Overall, the result of the experiment demonstrates that both Facet and `EplusM` are effective for visualizing OMVs, offering comparable or lower error rates ($\mathbf{H_{err}}$), similar or faster response times ($\mathbf{H_{tim}}$), and equal or higher confidence levels ($\mathbf{H_{con}}$), compared with Lin, Log, and SSB. These findings suggest that the visualizations identified as "Effective assignments" in our qualitative evaluation, which adhere to the proposed design guidelines, not only match but may surpass the performance of log scale bar charts and state-of-the-art solutions like SSB across key effectiveness metrics. This advantage is particularly pronounced for quantitative comparisons involving values of similar and neighboring magnitudes.

Our results align with previous studies that used the same error measurement [30], where linear, logarithmic, and SSB charts showed similar performance in terms of time and accuracy. However, they differ from experiments that employed different error measures [29], where SSB significantly outperformed Log for ratio estimation tasks. This discrepancy may be attributed to the use of Log Relative Error, which, while advantageous over Absolute Relative Error in some contexts [70], appears to favor logarithmic scale evaluation. For instance, a 1-digit mantissa error translates to 0.3 between values of 1 and 2 but drops to just 0.05 (six times smaller) between 8 and 9. This suggests that Log Relative Error benefits logarithmic scales, as errors at higher values— where ticks are denser and retrieving values is more difficult—are weighted less than equivalent errors at lower values within the same interval between two exponents. However, to our knowledge, there is no proposed relative error measure in the literature that addresses this bias and is suitable for analyzing data with high heteroscedasticity. Despite this, we believe the choice of error measure did not affect the comparison between our proposed designs and SSB, as all Facet, `EplusM`, and SSB use linear scales between the orders of magnitude.

Furthermore, our results demonstrate that it is essential to evaluate the effectiveness of visualizations per task, as both the task and the magnitude difference between compared values influence performance. For instance, in value retrieval tasks, we didn't observe important difference in accuracy between the evaluated designs. However, the effect size was more pronounced in ratio estimation tasks, particularly for values within the same or neighboring magnitudes. Participants reported difficulties with logarithmic scales in these tasks, as reflected in free-text feedback, where one participant noted, *"Since the bars don't show the actual value in a more ordinate way, it was hard for me to calculate the ratio and be more precise in some cases where the jump was bigger."* Interestingly, although linear bar charts did not perform well in most tasks, participants were able to estimate differences between values in distant magnitudes. This can be attributed to the high probability of having at least one bar within the visible range of linear scales (three larger exponents) in this task. This visibility allowed participants to estimate the difference by focusing on the value of the visible bar, as the difference between distant values was considered negligible.

7 GENERAL DISCUSSION, LIMITATIONS, AND FUTURE WORK

The objective of our work is to effectively visualize OMVs in overview, inspired by the natural human tendency to conceptualize large numbers as a combination of two components, such as mantissa and exponent. Unlike many design space explorations in visualization literature [13, 32, 62] that build on existing designs, we adopted a different systematic approach, which we believe can be generalized to other scenarios. Our exploration was grounded in the principles of the Grammar of Graphics, which provided a structured foundation for specifying the dimensions of our design space and facilitated the systematic generation of a wide range of visualizations. We then incorporated insights from numerical and graphical perception studies to derive design requirements and constraints, refining the space to a set of viable solutions. These solutions were generated and evaluated using an inspection-based qualitative method to identify potential encoding and decoding challenges, ultimately leading to the development of a set of design guidelines.

Among the solutions, collinear positional encodings for both mantissa and exponent (`EplusM` and Facet) emerged as the most effective. They provided highly accurate and discriminable channels for both the exponent (Accuracy for Magnitudes) and mantissa (Detail inside Magnitudes), preserved the natural continuity between them (Continuity Between Magnitudes), and supported the integration of additional data attributes while minimizing interference effects (Parsimony in Channels). A crowdsourcing evaluation demonstrated that `EplusM` and Facet performed as well as, or better than, existing solutions for specific tasks, validating our proposed guidelines and reinforcing the importance of separating OMVs into mantissa and exponent for accurate quantitative comparisons. These contributions advance the visualization literature on OMVs and enhance the state of the art.

Previous studies have already proposed two collinear positional designs, SSB and OML, which align closely with our design guidelines and bear similarities to the Facet and `EplusM` visualizations that emerged from our exploration. Both SSB and Facet encode the exponent using rows; however, they differ in how they handle the mantissa. SSB resets the range within each row starting from 0, whereas Facet encodes the mantissa in PosY within the range $[1, 10)$. While SSB 's approach supports across-magnitude comparisons by enabling users to interpret rows independently, it disrupts continuity between magnitudes and increases visual clutter—factors likely contributing to its higher response times in most tasks. The positional scales of the OML and `EplusM` designs differ primarily in how they visually represent the mantissa and exponent along PosY. OML separates these components, requiring users to multiply them, while `EplusM` integrates both into a single value. This subtle difference could facilitate users' interpretation, as they do not need to understand the concept of mantissa and exponent or perform mental calculations to determine the final value. Considering the visual similarities with prior work, we do not claim novelty for the `EplusM` and Facet designs, but regard them as refined versions and effective representations of our design guidelines.

While our study addressed key questions regarding OMV visualizations, it also raised new ones and was unable to resolve all challenges. We did not attempt to explore the full design space of possible OMV visualization designs, and thus our proposed design space is consistent but not complete. By focusing on a simplified subset—tabular data, Cartesian layouts, single-mark visualizations, and no redundant encoding—we were able to systematically investigate perceptual and cognitive challenges related to separating OMVs into mantissa and exponent. This choice served as an essential first step in understanding the visual interplay between these two components before expanding into a broader design space. Although our guidelines may not represent a complete set of rules for effective OMV visualizations, we believe they offer a useful framework for designers and researchers to explain their design rationale and identify potential encoding and decoding issues.

Future research should build on these findings by exploring a wider variety of data types, incorporating redundant encoding or interactivity, and assessing additional tasks. Future studies should also expand evaluations to more complex tasks, such as identifying correlations or supporting decision-making. Additionally, this study focused primarily on the perceptual and cognitive aspects of visualization, leaving out higher-level considerations such as memorability, trust, aesthetics, and user engagement. Future research should evaluate the real-world applicability and differences between `EplusM` and Facet, particularly in their effectiveness at communicating OMVs to a general audience without prior training. We believe that our study results generalize to other domains, and mantissa-exponent visualizations could prove beneficial for networks, fields, maps, and specialized visualization applications.

We hope our work inspires visualization designers and researchers to develop novel and effective OMVs designs, following and extending our proposed guidelines. We also encourage system developers to implement operators that can easily separate OMVs into mantissa and exponent components. We envision support for OMVs in popular visualization systems, requiring minor adjustments: (1) introducing the `EplusM` scale, (2) providing functions to extract exponent and mantissa parts of a value , and (3) introducing a new quantitative type (e.g., `omv`) in addition to `quantitative`, allowing for automatic use of the `EplusM` scale when appropriate and enabling syntax for accessing the individual parts.

8 CONCLUSION

This work explores the design of mantissa and exponent visualizations for OMVs. We proposed a design space focused on tabular data, informed and constrained by graphical perception principles. To better understand the interplay between mantissa and exponent, we qualitatively evaluated candidate visualizations within this space, identifying encoding and decoding challenges that could impact their effectiveness. Based on these evaluations, we derived design guidelines and validated them through a quantitative comparison of two visualizations aligned with these guidelines against traditional linear and logarithmic bar charts, as well as a state-of-the-art visualization. Our results demonstrate that adhering to our proposed guidelines and separating OMVs into mantissa and exponent components improves performance in quantitative comparisons. Floating-point numbers made possible calculations with OMVs on computers; scientific notation allows human reading of OMVs; our work contributes to generalizing the use of mantissa and exponent visualizations for OMVs.

ACKNOWLEDGMENTS

The authors wish to thank Ludovic David for his support in the development and bulk generation of images, Olivier Gladin for his support with the wall-size display, Lorenzo Ciccione, and Emanuele Santos for their useful feedback and thoughtful comments. This work has been partially supported by Berger-Levrault.

SUPPLEMENTAL MATERIALS

All supplemental materials are available on OSF at https://osf.io/uke76/?view_only=2ab48fcf974e47a2a9d1f7226e1b3c8b, released under a CC BY 4.0 license. In particular, they include (1) the dataset we used for our qualitative exploration, (2) the `combinations.js` file that we used to apply the constraints and generate the viable combinations, (3) a pdf file providing more details on the implementation of our tool, showcasing usage examples, and including a link to the code repository, (4) a website with the 168 visualizations that are included in our design space, also accessible via: https://www.omvs-designspace.com/, (5) csv files with the results of our qualitative evaluation, (6) the website developed for the crowdsource evaluation, (7) the anonymized experimental data, (8) the Jupiter notebooks used for the analysis, (9) the tables with the results, (10) the implementations of `EplusM` and Facet.

REFERENCES

[1] Daniel Algom and Daniel Fitousi. 2016. Half a century of research on Garner interference and the separability-integrality distinction. *Psychol. Bull.* 142, 12 (Dec. 2016), 1352–1383. https://doi.org/10.1037/bul0000072

[2] Emily Ayubi (Ed.). 2020. *Publication Manual of the American Psychological Association (7th Ed.).* APA, Washington. https://doi.org/10.1037/0000165-000

[3] Michael Behrisch, Michael Blumenschein, Nam Wook Kim, Lin Shao, Mennatallah El-Assady, Johannes Fuchs, Daniel Seebacher, Alexandra Diehl, Ulrik Brandes, Hanspeter Pfister, Tobias Schreck, Daniel Weiskopf, and Daniel A. Keim. 2018. Quality Metrics for Information Visualization. *Comput. Graph. Forum* 37, 3 (June 2018), 625–662. https://doi.org/10.1111/CGF.13446

[4] Jacques Bertin. 1983. *Semiology of graphics.* University of Wisconsin Press, USA.

[5] Michael Mose Biskjaer, Peter Dalsgaard, and Kim Halskov. 2014. A Constraint-Based Understanding of Design Spaces. In *Proceedings of the Conference on Designing Interactive Systems (DIS).* ACM, USA, 453–462. https://doi.org/10.1145/2598510.2598533

[6] Rita Borgo, Joel Dearden, and Mark W. Jones. 2014. Order of Magnitude Markers: An Empirical Study on Large Magnitude Number Detection. *IEEE Trans. Vis. Comput. Graph.* 20, 12 (Dec. 2014), 2261–2270. https://doi.org/10.1109/TVCG.2014.2346428

[7] Rita Borgo, Johannes Kehrer, David H. S. Chung, Eamonn Maguire, Robert S. Laramee, Helwig Hauser, Matthew Ward, and Min Chen. 2013. Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications. In *Proceedings of the Annual Conference of the European Association for Computer Graphics (Eurographics).* Eurographics Association, Girona, Spain, 39–63. https://doi.org/10.2312/CONF/EG2013/STARS/039-063

[8] Rita Borgo, Luana Micallef, Benjamin Bach, Fintan McGee, and Bongshin Lee. 2018. Information Visualization Evaluation Using Crowdsourcing. *Comput. Graph. Forum* 37, 3 (June 2018), 573–595. https://doi.org/10.1111/CGF.13444

[9] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D^3 Data-Driven Documents. *IEEE Trans. Vis. Comput. Graph.* 17, 12 (2011), 2301–2309. https://doi.org/10.1109/TVCG.2011.185

[10] Daniel Braun, Rita Borgo, Max Sondag, and Tatiana von Landesberger. 2024. Reclaiming the Horizon: Novel Visualization Designs for Time-Series Data with Large Value Ranges. *IEEE Trans. Vis. Comput. Graph.* 30, 1 (2024), 1161–1171. https://doi.org/10.1109/TVCG.2023.3326576

[11] Daniel Braun, Kerstin Ebell, Vera Schemann, Laura Pelchmann, Susanne Crewell, Rita Borgo, and Tatiana von Landesberger. 2022. Color Coding of Large Value Ranges Applied to Meteorological Data. In *Visualization and Visual Analytics (VIS).* IEEE, USA, 125–129. https://doi.org/10.1109/VIS54862.2022.00034

[12] Virginia Braun and Victoria Clarke. 2021. *Thematic Analysis: A Practical Guide.* SAGE, London.

[13] Matthew Brehmer, Bongshin Lee, Benjamin Bach, Nathalie Henry Riche, and Tamara Munzner. 2017. Timelines Revisited: A Design Space and Considerations for Expressive Storytelling. *IEEE Trans. Vis. Comput. Graph.* 23, 9 (2017), 2151–2164. https://doi.org/10.1109/TVCG.2016.2614803

[14] Antoine Béland, Florent Daudens, and Thomas Hurtut. 2023. Waffster: Hierarchical Waffle Charts for Budget Visualization. *Information Visualization* 22, 4 (Oct. 2023), 279–290. https://doi.org/10.1177/14738716231173730

[15] Tara C. Callaghan. 1984. Dimensional interaction of hue and brightness in preattentive field segregation. *Percept. & Psychophys.* 36, 1 (Jan. 1984), 25–34. https://doi.org/10.3758/BF03206351

[16] Gary Charness, Uri Gneezy, and Michael A. Kuhn. 2012. Experimental Methods: Between-Subject and Within-Subject Design. *J. Econ. Behav. Organ.* 81, 1 (Jan. 2012), 1–8. https://doi.org/10.1016/j.jebo.2011.08.009

[17] Lorenzo Ciccione, Mathias Sablé-Meyer, and Stanislas Dehaene. 2022. Analyzing the Misperception of Exponential Growth in Graphs. *Cogn.* 225 (Aug. 2022), 105112:1–105112:11. https://doi.org/10.1016/j.cognition.2022.105112

[18] William S. Cleveland and Robert McGill. 1984. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *J. Am. Stat. Assoc.* 79, 387 (1984), 531–554. https://doi.org/10.1080/01621459.1984.10478080

[19] Cristina Conati, Giuseppe. Carenini, Enamul Hoque, Ben Steichen, and Dereck Toker. 2014. Evaluating the Impact of User Characteristics and Different Layouts on an Interactive Visualization for Decision Making. *Comput. Graph. Forum* 33, 3 (June 2014), 371–380. https://doi.org/10.1111/cgf.12393

[20] Pierre Dragicevic. 2016. Fair Statistical Communication in HCI. In *Modern Statistical Methods for HCI.* Springer, USA, 291–330. https://doi.org/10.1007/978-3-319-26633-6_13

[21] Ben H. Engler, Florian Hutzler, and Stefan Hawelka. 2024. An extended logarithmic visualization improves forecasting accuracy for exponentially growing numbers, but residual difficulties remain. *International Journal of Forecasting* (2024), to appear. https://doi.org/10.1016/j.ijforecast.2024.09.006

[22] Franz Faul, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang. 2009. Statistical Power Analyses Using G*Power 3.1: Tests for Correlation and Regression Analyses. *Behav. Res. Methods* 41, 4 (Nov. 2009), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

[23] Tzvi Ganel and Melvyn A. Goodale. 2003. Visual Control of Action but not Perception Requires Analytical Processing of Object Shape. *Nature* 426, 6967 (Dec. 2003), 664–667. https://doi.org/10.1038/nature02156

[24] Wendell R. Garner. 1976. Interaction of Stimulus Dimensions in Concept and Choice Processes. *Cogn. Neuropsychol.* 8, 1 (Jan. 1976), 98–123. https://doi.org/10.1016/0010-0285(76)90006-2

[25] David Goldberg. 1991. What Every Computer Scientist Should Know About Floating-Point Arithmetic. *ACM Comput. Surv.* 23, 1 (March 1991), 5–48. https://doi.org/10.1145/103162.103163

[26] Richard L. Gottwald and Wendell R. Garner. 1975. Filtering and condensation tasks with integral and separable dimensions. *Percept. Psychophys.* 18, 1 (Jan. 1975), 26–28. https://doi.org/10.3758/BF03199362

[27] Christopher G. Healey and James T. Enns. 2012. Attention and Visual Memory in Visualization and Computer Graphics. *IEEE Trans. Vis. Comput. Graph.* 18, 7 (July 2012), 1170–1188. https://doi.org/10.1109/TVCG.2011.127

[28] Jeffrey Heer and Michael Bostock. 2010. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. In *Proceedings of the International Conference on Human Factors (CHI).* ACM, USA, 203–212. https://doi.org/10.1145/1753326.1753357

[29] Marcel Hlawatsch, Filip Sadlo, Michael Burch, and Daniel Weiskopf. 2013. Scale-Stack Bar Charts. *Comput. Graph. Forum* 32, 3pt2 (June 2013), 181–190. https://doi.org/10.1111/cgf.12105

[30] Markus Höhn, Marcel Wunderlich, Kathrin Ballweg, and Tatiana von Landesberger. 2020. Width-Scale Bar Charts for Data with Large Value Range. In *Proceeding of the Eurographics Conference on Visualization (EuroVis).* Eurographics Association, Norrköping, Sweden, 103–107. https://doi.org/10.2312/EVS.20201056

[31] Petra Isenberg, Anastasia Bezerianos, Pierre Dragicevic, and Jean-Daniel Fekete. 2011. A Study on Dual-Scale Data Charts. *IEEE Trans. Vis. Comput. Graph.* 17, 12 (Dec. 2011), 2469–2478. https://doi.org/10.1109/TVCG.2011.160

[32] Waqas Javed and Niklas Elmqvist. 2012. Exploring the design space of composite visualization. In *2012 IEEE Pacific Visualization Symposium, PacificVis 2012, Songdo, Korea (South), February 28 - March 2, 2012,* Helwig Hauser, Stephen G. Kobourov, and Huamin Qu (Eds.). IEEE, 1–8. https://doi.org/10.1109/PACIFICVIS.2012.6183556

[33] Kaggle. 2024. Youtube trends. https://www.kaggle.com/datasets/datasnaek/youtube-new?resource=download&select=FRvideos.csv. Accessed on 22-Feb-2024.

[34] Alireza Karduni, Ryan Wesslen, Isaac Cho, and Wenwen Dou. 2020. Du Bois Wrapped Bar Chart: VisualizingCategorical Data with Disproportionate Values. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI).* ACM, USA, 1––12. https://doi.org/10.1145/3313831.3376365

[35] Younghoon Kim and Jeffrey Heer. 2018. Assessing Effects of Task and Data Distribution on the Effectiveness of Visual Encodings. *Comput. Graph. Forum* 37, 3 (June 2018), 157–167. https://doi.org/10.1111/CGF.13409

[36] Donald E. Knuth. 1997. *The Art of Computer Programming, Volume 2 (3rd ed.): seminumerical algorithms*. Addison-Wesley Longman Publishing Co., Inc., USA.

[37] David H. Krantz and Amos Tversky. 1975. Similarity of Rectangles: An Analysis of Subjective Dimensions. *J. Math. Psychol.* 12, 1 (Feb. 1975), 4–34. https://doi.org/10.1016/0022-2496(75)90047-4

[38] Richard M. Kubina, Seth A. King, Madeline Halkowski, Shawn Quigley, and Tracy Kettering. 2023. Slope Identification and Decision Making: A Comparison of Linear and Ratio Graphs. *Behavior Modification* 47, 3 (May 2023), 615–643. https://doi.org/10.1177/01454455221130002

[39] David Landy, Arthur Charlesworth, and Erin Ottmar. 2017. Categories of Large Numbers in Line Estimation. *Cognitive Science* 41, 2 (2017), 326–353. https://doi.org/10.1111/cogs.12342

[40] David Landy, Noah Silbert, and Aleah Goldin. 2013. Estimating Large Numbers. *Cognitive Science* 37, 5 (July 2013), 775–799. https://doi.org/10.1111/cogs.12028

[41] Jing Li, Jarke J. van Wijk, and Jean-Bernard Martens. 2010. A Model of Symbol Lightness Discrimination in Sparse Scatterplots. In *Proceedins of the Pacific Visualization Symposium (PacificVIS)*. IEEE, USA, 105–112. https://doi.org/10.1109/PACIFICVIS.2010.5429604

[42] Jock Mackinlay. 1986. Automating the Design of Graphical Presentations of Relational Information. *ACM Trans. Graph.* 5, 2 (April 1986), 110—-141. https://doi.org/10.1145/22949.22950

[43] Neil A. Macmillan and Adam S. Ornstein. 1998. The Mean-Integral Representation of Rectangles. *Percept. Psychophys.* 60, 2 (Jan. 1998), 250–262. https://doi.org/10.3758/BF03206034

[44] Jennifer Mankoff, Anind K. Dey, Gary Hsieh, Julie Kientz, Scott Lederer, and Morgan Ames. 2003. Heuristic Evaluation of Ambient Displays. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. ACM, USA, 169–176. https://doi.org/10.1145/642611.642642

[45] Caitlyn M. McColeman, Fumeng Yang, Timothy F. Brady, and Steven Franconeri. 2022. Rethinking the Ranks of Visual Channels. *IEEE Trans. Vis. Comput. Graph.* 28, 01 (Jan. 2022), 707–717. https://doi.org/10.1109/TVCG.2021.3114684

[46] Candace McNaughton, Kenneth A Wallston, Russell L Rothman, David E Marcovitz, and Alan B Storrow. 2011. Short, Subjective Measures of Numeracy and General Health Literacy in an Adult Emergency Department. *Acad Emerg Med* 18, 11 (Nov. 2011), 1148–1155. https://doi.org/10.1111/j.1553-2712.2011.01210.x

[47] Candace D McNaughton, Kerri L Cavanaugh, Sunil Kripalani, Russell L Rothman, and Kenneth A Wallston. 2015. Validation of a Short, 3-Item Version of the Subjective Numeracy Scale. *Med Decis Making* 35, 8 (April 2015), 932–936. https://doi.org/10.1177/0272989X15581800

[48] Duncan N. L. Menge, Anna C. MacPherson, Thomas A. Bytnerowicz, Andrew W. Quebbeman, Naomi B. Schwartz, Benton N. Taylor, and Amelia A. Wolf. 2018. Logarithmic scales in ecological data presentation may cause misinterpretation. *Nature Ecology & Evolution* 2, 9 (2018), 1393–1402. https://doi.org/10.1038/s41559-018-0610-7

[49] Miro. 2024. Miro. https://miro.com/. Accessed on 08-Sep-2024.

[50] Dominik Moritz, Chenglong Wang, Greg L. Nelson, Halden Lin, Adam M. Smith, Bill Howe, and Jeffrey Heer. 2019. Formalizing Visualization Design Knowledge as Constraints: Actionable and Extensible Models in Draco. *IEEE Trans. Vis. Comput. Graph.* 25, 1 (Jan. 2019), 438–448. https://doi.org/10.1109/TVCG.2018.2865240

[51] Tamara Munzner. 2014. *Visualization Analysis and Design*. CRC Press, Boca Raton, Florida, United States. https://doi.org/10.1201/b17511

[52] Plot. 2024. Observable Plot. https://observablehq.com/plot/. Accessed on 02-Feb-2024.

[53] Prolific. 2024. Prolific. https://https://www.prolific.com/. Accessed on 08-Sep-2024.

[54] Ghulam Jilani Quadri and Paul Rosen. 2022. A Survey of Perception-Based Visualization Studies by Task. *IEEE Trans. Vis. Comput. Graph.* 28, 12 (Dec. 2022), 5026–5048. https://doi.org/10.1109/TVCG.2021.3098240

[55] React. 2024. React. https://react.dev/. Accessed on 08-Sep-2024.

[56] Ronald A. Rensink. 2014. On the Prospects for a Science of Visualization. In *Handbook of Human Centric Visualization*, Weidong Huang (Ed.). Springer, USA, 147–175. https://doi.org/10.1007/978-1-4614-7485-2_6

[57] Kevin A. Richards and Michael A. Hemphill. 2018. A Practical Guide to Collaborative Qualitative Data Analysis. *J. Teach. Phys. Educ.* 37, 2 (April 2018), 225–231. https://doi.org/10.1123/jtpe.2017-0084

[58] Alessandro Romano, Chiara Sotis, Goran Dominioni, and Sebastián Guidi. 2020. The Scale of COVID-19 Graphs Affects Understanding, Attitudes, and Policy Preferences. *Health Econ.* 29, 11 (Nov. 2020), 1482–1494. https://doi.org/10.1002/hec.4143

[59] William Ryan and Ellen Evers. 2020. Logarithmic Axis Graphs Distort Lay Judgment. https://doi.org/10.2139/ssrn.3605872 (preprint).

[60] Sass. 2024. Syntactically Awesome Style Sheets. https://sass-lang.com/. Accessed on 08-Sep-2024.

[61] Arvind Satyanarayan, Dominik Moritz, Kanit Wongsuphasawat, and Jeffrey Heer. 2017. Vega-Lite: A Grammar of Interactive Graphics. *IEEE Trans. Vis. Comput. Graph.* 23, 1 (Jan. 2017), 341–350. https://doi.org/10.1109/TVCG.2016.2599030

[62] Hans-Jörg Schulz, Steffen Hadlak, and Heidrun Schumann. 2011. The Design Space of Implicit Hierarchy Visualization: A Survey. *IEEE Trans. Vis. Comput. Graph.* 17, 4 (2011), 393–411. https://doi.org/10.1109/TVCG.2010.79

[63] Andrew Sears. 1997. Heuristic Walkthroughs: Finding the Problems Without the Noise. *Int. J. Hum.–Comput. Interact.* 9, 3 (1997), 213–234. https://doi.org/10.1207/s15327590ijhc0903_2

[64] Herbert A. Simon. 1973. The Structure of Ill Structured Problems. *Artif. Intell.* 4, 3 (1973), 181–201. https://doi.org/10.1016/0004-3702(73)90011-8

[65] Stephen Smart and Danielle Albers Szafir. 2019. Measuring the Separability of Shape, Size, and Color in Scatterplots. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. ACM, USA, 669:1—-669:14. https://doi.org/10.1145/3290605.3300899

[66] Laura South, David Saffo, Olga Vitek, Cody Dunne, and Michelle A. Borkin. 2022. Effective Use of Likert Scales in Visualization Evaluations: A Systematic Review. *Comput. Graph. Forum* 41, 3 (June 2022), 43–55. https://doi.org/10.1111/CGF.14521

[67] Danielle Albers Szafir. 2018. Modeling Color Difference for Visualization Design. *IEEE Trans. Vis. and Comput. Graph.* 24, 1 (Jan. 2018), 392–401. https://doi.org/10.1109/TVCG.2017.2744359

[68] Michal Tanzer, Erez Freud, Tzvi Ganel, and Galia Avidan. 2013. General Holistic Impairment in Congenital Prosopagnosia: Evidence from Garner's Speeded-Classification Task. *Cogn. Neuropsychol.* 30, 6 (2013), 429–445. https://doi.org/10.1080/02643294.2013.873715

[69] Alvin Tarrell, Ann L. Fruhling, Rita Borgo, Camilla Forsell, Georges G. Grinstein, and Jean Scholtz. 2014. Toward Visualization-Specific Heuristic Evaluation. In *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization (BELIV)*. ACM, USA, 110–117. https://doi.org/10.1145/2669557.2669580

[70] Chris Tofallis. 2015. A Better Measure of Relative Prediction Accuracy for Model Selection and Model Estimation. *J. Oper. Res. Soc.* 66, 8 (Aug. 2015), 1352–1362. https://doi.org/10.1057/jors.2014.103

[71] Leo Törnqvist, Pentti Vartia, and Yrjö O. Vartia. 1985. How Should Relative Changes Be Measured? *Am. Stat.* 39, 1 (1985), 43–46. https://doi.org/10.1080/00031305.1985.10479385

[72] Melanie Tory and Torsten Möller. 2005. Evaluating Visualizations: Do Expert Reviews Work? *IEEE Comput. Graph. Appl.* 25, 5 (2005), 8–11. https://doi.org/10.1109/MCG.2005.102

[73] Hadley Wickham. 2010. A Layered Grammar of Graphics. *Journal of Computational and Graphical Statistics* 19, 1 (2010), 3–28. https://doi.org/10.1198/jcgs.2009.07098

[74] Leland Wilkinson. 2005. *The Grammar of Graphics*. Springer, USA. https://doi.org/10.1007/0-387-28695-0

[75] Zehua Zeng and Leilani Battle. 2023. A Review and Collation of Graphical Perception Knowledge for Visualization Recommendation. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. ACM, USA, 820:1–820:16. https://doi.org/10.1145/3544548.3581349

[76] Torre Zuk and Sheelagh Carpendale. 2006. Theoretical Analysis of Uncertainty Visualizations. In *Proceedings of Visualization and Data Analysis (VDA)*. SPIE, USA, 606007. https://doi.org/10.1117/12.643631